A Process for Developing Accurate Kinesic Cues in Virtual Environments

Stephanie J. Lackey
Karla A. Badillo-Urquiola
Eric C. Ortiz
Institute for Simulation and Training
University of Central Florida
Orlando, FL 32826
slackey@ist.ucf.edu, kbadillo@ist.ucf.edu, eortiz@ist.ucf.edu

Irwin L. Hudson
U.S. Army Research Laboratory
The Human Research and Engineering Directorate
Simulation and Training Technology Center
Orlando, FL 32826
irwin.hudson@us.army.mil

Keywords:
Behavior Cue Analysis, Kinesic Cues, Simulation-Based Training, Computer Animation

ABSTRACT: Computer animations exhibit the illusion of movements or actions of virtual agents and assets within a virtual environment display. Two distinct animation categories exist: two-dimensional (2D) and three-dimensional (3D). 2D animation is typically stylized and used primarily for entertainment-based efforts such as cartoons and low-fidelity games. 3D animation is applied to a wider variety of domains (e.g., entertainment games, serious games, and training simulations). A well-designed 3D computer animation enables a realistic representation of action portraying the true context of movement, particularly human gestures (Badler, Palmer, & Bindiganavale, 1999). All humans convey intent whether purposefully or not via verbal and non-verbal cues (Bavelas, 1990; Givens, 2002). Kinesic cues convey information to an observer through body language and gestures. Emerging research in training human threat detection requires virtual agents exhibiting kinesic cues to provide visual stimuli within Simulation-Based Training (SBT) applications. Thus, guidelines and specifications for system developers are required. This paper presents a process for defining, designing, and animating kinesic cues using a commercially available software application to mimic realistic human behaviors, movements, and gestures. Through this discussion, culturally agnostic kinesic cues are presented, and relevant limitations are identified. The process described and lessons learned represent a logical progression in the formalization of developing advanced visual models for training Warfighters, law enforcement agents, and first responders to detect and classify human threats.

1. Introduction

Reading the human terrain is a critical skill for Warfighters, law enforcement, and first responders. Detection and classification of explicit and implicit behavioral cues relies on observation and perception of visual cues such as gestures, facial expressions, and posture. Furthermore, timely detection and classification of behavioral cues enables Warfighters to proactively identify potential threats within a given environment (Flynn, Sisco, & Ellis, 2012). In order to make proactive tactical decisions, understanding the typical day-to-day operations of an environment, or baseline, is required to efficiently and effectively identify anomalies (see Figure 1.1).

Combat Profiling serves as the foundation for much of the previous behavior cue analysis research, development, and training funded by the U.S. Armed Forces (Gideons, Padilla, & Lethin, 2008; Van Horne, 2013).
The core concepts of Combat Profiling are being extended to enhance training of active duty Warfighters, law enforcement officers, and civilians (Van Horne, 2013), and the associated training methods are evolving, as well. Historically, training such skills occurred in traditional classroom paradigms (Gideons, Padilla, & Lethin, 2008), but recent research and development efforts are moving behavior cue detection into SBT environments (Schatz, Wray, Folsom-Kovarik, & Nicholson, 2012; Wang-Costello, Tarr, & Marraffino, 2013). Improving the effectiveness and distribution of this type of training requires advancing both training methods and technologies. One of the remaining challenges faced by the SBT community is to define requirements for visually representing behavior cues in virtual environments.

This research aims to clearly articulate requirements for 3D animations of behavior cues in order to directly support ongoing research investigating the training efficacy of traditional SBT strategies (e.g., highlighting, massed exposure, etc.) and game-based principals (e.g., leader boards, tokens, etc.). Thus, a Behavior Cue Catalog (BCC) resulted from a multi-year effort to theoretically define, describe, categorize, and prototype 3D models of biometric and kinesic cues. For the 12 kinesic cues presented in this paper, the theoretically based definition, explanation of possible cause(s), and verbal and visual guidance documented in the Behavior Cue Catalogue is summarized. (For a detailed description of the biometric research found in the BCC see Lackey, Badillo-Urquiola, and Ortiz, 2014.)

Before discussing the research approach and results, it is important to provide background information about the field of Kinesics. Kinesics is the interpretation of non-verbal cues such as body language, gestures, postures, and facial expressions displayed by individuals (Birdwhistell, 1970; Leathers, 1997). Four key categories have been identified for classifying kinesic cues: manipulators, illustrators, regulators, and emblems (Ekman, 1992; Ekman, 2004; Leathers, 1997). Manipulators typically indicate an effort to soothe oneself. One part of the body massages, rubs, pinches, scratches, or manipulates another part of the body to relieve stress or discomfort. In some circumstances, manipulators can also indicate a state of relaxation. Illustrators offer additional context to other modes of communication. This type of cue may be confused with emblems; however, illustrators facilitate the listener’s understanding of messages (Ekman, 1992). Illustrators are used to emphasize a word or phrase and when something is difficult to explain; however, they have no meaning separate from context of speech (Harrington, 2009). Behavioral cues that monitor or control the listener of a conversation are known as regulators. Regulators serve as conversational mediators (Ekman, 2004). Finally, emblems represent culturally specific cues understood by every individual in a culture group and do not need words to have meaning (Ekman, 1992). Illustrators may be confused with emblems (i.e., culturally specific cues); however, they facilitate the listener’s understanding of messages (Ekman, 1992). Although illustrators typically emphasize a word or phrase when something is difficult to explain, they have no meaning separate from context of speech (Harrington, 2009). This paper focuses on universal behavior cues (Navarro & Karlins, 2008), thus examples of emblems will not be addressed in detail, but the concept is listed here for completeness.

2. Approach

A systematic, four step approach was used to create the kinesic models depicted in the BCC. The first step was to document design specifications based upon a literature review and Subject Matter Experts (SME) input. Second, a 3D artist created sample models depicting the cues defined. Next, preliminary validity testing confirmed adherence to specifications or identified areas of improvement for each model. Finally, the models and design documentation were modified per the test results. The next section provides relevant details from each phase.

2.1 Define Requirements and Design Specifications

The Non-Verbal Dictionary of Gestures, Signs, and Body Language (Givens, 2002), served as the foundation for understanding kinesic cues, and was augmented by an extensive literature review that referenced a variety of disciplines including: psychology, sociology, communications, criminal justice, political science, health, and medical professions. Historic and contemporary scholarly sources were consulted including books, journal articles, and conference papers, in addition to insight gained from SMEs. Four cue categories emerged: manipulators, regulators, illustrators, and emblems. As previously stated, the first three categories are the focus of the research presented.

Upon completion of the identification and categorization phase, the research team consulted with SMEs to prioritize the cues addressed. Following cue prioritization, the research team defined each cue and developed illustrative descriptions to facilitate 3D modeling efforts. Resulting design specifications include concise definitions culled from the literature, explanations of possible underlying causes, and implications for creating virtual models. The cues included in this paper represent cues that were identified in the literature, prioritized by SMEs, and
could be prototyped using typical commercially available tools. Three cues investigated, but excluded, are described in the limitations section.

2.2 Develop Models

The development phase followed. All 3D models utilized for this effort were purchased and downloaded from a third party digital media reseller. Design parameters (e.g., high-polygon count, high-resolution UV-mapped textures, rigging, and rendering attributes) served as selection criteria. In order to fully understand the animation process employed, a few key terms require explanation.

High-Polygon Count: Polygon count refers to the number of polygons used to create a model. Polygons in 3D modeling are simple 2D shapes enclosed and connected to comprise a full model. The more polygons a model contains, the more realistic its appearance in both static and dynamic imagery. Models with high-polygon counts, considered high-fidelity, increase demand on computer processing power in 3D space and Virtual Environments (VE) depending on the number of virtual agents. For the purpose of this effort, the polygon counts for the models utilized averaged between 7,500 and 10,000, and processing power presented no limitations.

Ultraviolet (UV)-Mapped Textures: Each model included pre-designed, high-resolution, UV-mapped textures to accurately portray photorealistic human facial features, skin tones, and clothing. Basic wire frames constitute the foundation of all models prior to defining the surface appearance by applying 2D texture maps to UV coordinates. 3D software applications use horizontal and vertical axes called UV coordinates to project 2D images onto the model. Texture maps depict objects such as faces, clothing, or hair. Typically, multiple detailed texture maps shape the appearance of each model. A texture map’s level of detail directly correlates to the precision of a model’s appearance. Photorealistic texturing aids in making a model of a human seem more lifelike.

Rigging: Rigging fits a solid model with a skeletal system, or “bones,” and manipulators for controlling and posing the model during the animation phase. The number of bones included in a human model varies according to the needs, specifications, and final use of the model. The rigging process defines how the bones shape and move a model. The level of detail required to appropriately represent a particular movement drives the complexity of the rigging process. Model objects called manipulators and controls drive motion and the number of manipulators and controls vary according to the complexity of the movements displayed. Animators use various references such as motion-capture, video footage, or live actors to ensure animations to properly convey realistic actions, timing, and gestures.

Key Framing: An industry standard for replicating motion in computer animations, key framing, depicts movement or motion according to start and stop points on a timeline. This process plays a key role in portraying realistic, life-like human movements. Smoothly transitioning from one position/movement to another involves multiple key frame sequences. Each movement possesses its own set of key frames. Each part of the model being animated may have interacting transitions such as raising an arm from a downward position, rotating the wrist, and placing it behind the head. Multiple key frames enable realistic lifelike movements.

Rendering: Rendering represents the final step in displaying 3D models and animations on a computer, and refers to the process by which the computer converts models and attributes in a 3D environment into 2D images. Rendering can include individual frames, sequences of frames, or consecutive sequences (i.e., an animation). Different rendering algorithms produce varying results and range from cartoon-like to photorealistic high-fidelity imagery.

Export: Following the rendering process, an animation must be exported from the 3D modeling application in order to be imported into the final medium (e.g., game engine or digital movie). Not all 3D software applications use output formats native to the final application. However, third party conversion applications enable streamlined conversions.

Animation process: Developing 3D computer animations that accurately portrayed human behaviors, movements, and gestures detailed in the Behavior Cue Catalog required a systematic approach. The following steps summarize the process used by the research team to animate the specified models:

1. Identified individual models possessing the necessary parameters (e.g., high-polygon count, high-resolution UV-mapped textures, rigging, and rendering attributes) and imported models into a commercially available 3D software application (i.e., Autodesk’s 3ds Max).

2. Scaled models in the 3D environment to maintain a standard, uniform size. The 3D software application utilized represented measurements in units. All 3D models were scaled proportionately to a 1 unit = 1 foot ratio. Each male 3D model represented a
human height of 6'0" and each female 3D model represented a human height of 5'7".

3. Reviewed and test rendered all texture maps to ensure objects associated with each model displayed properly.

4. Observed, at minimum, two males and two females of varying skin tones demonstrating each cue. These visualizations were video recorded and used as references by an artist to animate and replicate each pose on the 3D models.

5. Key framed each model’s motion on a timeline by moving and shifting the rigging manipulators in the 3D space to represent real-world movements. This process was the most time consuming in developing the animations because of the criticality of accurately animating subtle nuances associated with certain gestures (e.g., covering the mouth or hands on hips). The animator replayed each animation several times to ensure each action conveyed the intended message (within the bounds of the technology capabilities and limitations).

6. Rendered each model animation. The rendered output files were produced by the application playing the final animation (e.g., game engine or digital movie). The final renders for this effort were both single frames and animations. The single frames were rendered as high-resolution jpeg images and the animation sequences were Windows movie files. The render sizes or resolution was based upon the final output sizes needed per animation (e.g., whole body or facial features only).

2.3 Preliminary Validity Test and Model Revision

The members of the requirements and specification team analyzed the visual representation of each cue and assessed the accuracy of the behavior exhibited compared to the documented requirements. Discrepancies were noted and shared with the 3D artist to facilitate fine-tuning of each model. For instance, the fingers of the models exhibiting the wring hands cue moved more than expected. The amount of movement for the individual fingers was subsequently minimized as much as possible by the 3D artist. For example, the hands on the fair skinned female appeared excessively large compared to the relative size of the model’s body, and the fingernails on the medium skin toned male appeared unusually long after the first iteration of modeling. Each feature was adjusted until expected relative sizes were represented.

3. Results

The resulting models presented below provide sample still images of the animations developed. Up to four cues are presented for each cue type (i.e., manipulators, illustrators, and regulators). The results presented focus on the typical visual depiction of each cue as opposed to a specific scenario. In general, the first appearance of a cue or a decreased frequency in the display of a gesture can represent a low level of intensity. Alternatively, a high level of intensity for this effort could be defined as repeated or constant display of a cue. Such decisions depend on scenario details and context; and thus require definition in scenario design documentation.

3.1 Manipulators

Clenched Fists: The hand is closed, with fingers curled and squeezed into palms. This typically signals an emotional state of aggression or anger (Givens, 2002; Lackey, Maraj, Salcedo, Ortiz, & Hudson, 2013). The individual may be unconsciously displaying their anxiety (Givens, 2002). To relieve the stress, the individual clenches their fist. See Figure 3.1.

Hand Behind Head: This cue is exhibited by rubbing, scratching, or touching the back of the neck or head with an unclenched hand. Typical causes include a sense of uncertainty or discomfort, because the individual may be in disagreement or have frustrations (Givens, 2002). This gesture typically reflects negative emotions or thoughts (Givens, 2002). To relieve stress, the person touches or rubs the back of their head. Figure 3.2 below demonstrates this cue.
**Cover Mouth with Hand:** In this case, the individual places their hand in front or on top of the mouth. Someone telling a lie will cover their mouth indicating they are metaphorically trying to hide the lie. This action suggests that the liar is trying to hide themselves from the person they are trying to deceive (Kassin & Fong, 1999). Touching the lips with fingers also induces a sense of relaxation or stress relief (Givens, 2002). See Figure 3.3.

**Cover Eyes with Hand:** An individual will close or cover their eyes to metaphorically disregard the conversation or ignore the object in front of them. This cue can be considered a protective gesture because the hand is in contact with the face. Avoiding eye contact by closing or covering the eyes prevents anxiety (Blair & Kooi, 2003). See Figure 3.4.

**Wring Hands:** Squeezing and twisting both hands together suggest a desire to relieve stress and nervousness in an effort to find comfort (Navarro & Karlins, 2008). See Figure 3.5.

**Rub Palms:** Moving palms back and forth against each other typifies nervousness; rapid rubbing can indicate relaxation or expecting a successful result (Navarro & Karlins, 2008). Refer to Figure 3.6.

**3.2 Illustrators**

**Point Finger:** To extend forefinger (also referred to as index finger) in a stabbing motion denotes anger, aggression, hostility, or unfriendly action (Givens, 2002; Navarro & Karlins, 2008). Pointing without a stabbing motion is used to indicate direction or to direct someone’s attention (Givens, 2002). Refer to Figure 3.7.

**Slap Forehead:** When an individual hits his or her forehead with their hand, it typically indicates that the person is honestly recalling information and telling the truth. This can also be an indicator of uncertainty (Givens, 2002). See Figure 3.8.
3.3 Regulators

*Check Six:* The term Check Six, is short for the phrase “check your six o’clock” (Lackey, Maraj, Salcedo, Ortiz, & Hudson, 2013), may also be referred to as Cut-Off (Givens, 2002). In this case, an individual turns their head to look over their shoulder or turns their entire body around 180° (Lackey, Maraj, Salcedo, Ortiz, & Hudson, 2013). It depicts nervousness, uncertainty, or disagreement. If the individual is displaying a sustained cut-off, it may indicate shyness or disliking (Givens, 2002). See Figure 3.9 for an example.

*Figure 3.9 Slap Forehead*

*Figure 3.10 Check Six*

*Palms Down:* Expression of this cue involves one palm facing downward. Having the palms face downward indicates dominance, confidence, and assertiveness. Palms down can also be accompanied by aggressiveness (Givens, 2002).

*Palms Up:* This cue involves palms facing upward in an open position and indicates thoughtfulness or uncertainty (Givens, 2002).

*Hands on Hips:* Hands on Hips illustrates both phenomena described above. Palms placed onto the hips with elbows flexed outward, away from the body with palms facing downward depicts an assertive attitude (e.g., anger). See Figure 3.10. Alternatively, Figure 3.11 shows palms facing up due to the rotation of the thumbs forward (analogous to facing upward) (Givens, 2002).

*Figure 3.11 Hands on Hips Palms Down*

*Figure 3.12 Hands on Hips Palms Up*

4. Discussion

The results presented offer insight into communicating scenario design elements with simulation or model developers. The specifications and visualizations shared offer insight into how to incorporate such requirements into scenario design and test documentation. Ongoing experimentation will determine the efficacy of the kinesic cue specifications reported. However, preliminary experimental results suggest the method and procedure presented positively impact the ability to represent kinesic cues in VEs and SBT. Ultimately, the work presented provides a foundation upon which to build and illustrates the benefits of an interdisciplinary approach to addressing research gaps in the field of behavior cue training development.

5. Limitations

The lack of consensus in the literature concerning the underlying causes of body language and gestures presented a significant challenge. Consulting with SMEs from a relevant task domain, military operations, mitigated this risk. Context plays a critical role in behavior cue diagnosis and analysis. The literature related to culturally agnostic cues and the driving force behind them is limited. However, the research team carefully crafted the definitions presented by balancing tradeoffs between basic and applied research.

Of the cues presented, the Wring Hands cue posed the most difficulty to accurately simulate. Rotation constraints on the wrist and finger joints caused
asynchronous movement of the individual fingers. Minimization of this excessive movement partially addressed this issue but compromised the naturalness of the animation. Three additional cues proved too challenging to sufficiently animate due to technology limitations: Gaze Avoidance, Zygomatic Smile, and Cross Arms. These cues involve multiple, linked attributes or level of detail beyond the capability of standard commercial modeling applications. Research efforts are continuing to confront these complexity challenges.

Finally, the models presented require verification. At the time of writing, SME validation steered animation adjustments. Ongoing human participant experiments using the model specifications presented will serve to remedy this limitation.

6. Conclusion

The purpose of this research was to develop science-driven design requirements and recommendations for accurately modeling kinesic cues in VEs. A natural question that follows is, “Are these models valid?” Empirical validation of the 12 models presented is the first recommendation for future work. Similarly, the question “Are we modeling all of the appropriate cues?” warrants attention. Recommendations for future work include investigating the comprehensive list of cues categorized by domain (e.g., combat, law enforcement, self-defense). Furthermore, the integration of culturally agnostic and culturally specific cues would fuse disparate veins of research: universal behavior cue analysis and human social cultural behavior modeling. In addition to these theoretical research questions, the technical implementation requires attention from the applied research and industry perspective. Advancement of 3D animation capabilities, including detailed modeling of finger, hand, and facial motions, is recommended. Finally, moving this type of behavior representation beyond the contemporary standard in SBT platforms into Virtual Worlds opens up opportunities for large-scale training events. The initial behavior cue design specifications presented herein set the stage for improving Warfighter training in simulation-based environments.

7. Acknowledgements

This research was sponsored by the U.S. Army Research Laboratory – Human Research Engineering Directorate Simulation and Training Center (ARL HRED STTC), in collaboration with the Institute for Simulation and Training at the University of Central Florida. This work is supported in part by ARL HRED STTC contract W91CRB08D0015. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of ARL HRED STTC or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

8. References


Author Biographies

**STEPHANIE J. LACKEY** earned her Master’s and Ph.D. degrees in Industrial Engineering and Management Systems with a specialization in Simulation, Modeling, and Analysis at the University of Central Florida (UCF). Her research focused on prediction, allocation, and optimization techniques for digital and analog communications systems. She joined UCF Institute for Simulation and Training’s Applied Cognition and Training in Immersive Virtual Environments Lab in 2008, and assumed the role of Lab Director in 2010. Dr. Lackey leverages her experience in advanced predictive modeling to the field of human performance in order to develop methods for improving human performance in simulation-based training environments and human-robot interfaces.

**KARLA A. BADILLO-URQUIOLA** is a Graduate Research Assistant at the Applied Cognition and Training in Immersive Virtual Environments Lab. In January of 2012, she earned the UCF Undergraduate Researcher of the Month award. Her research interests include: simulation-based training; immersive virtual training environments; physiological & human factors; terrorism and counter-terrorism. As a recipient of the McNair Graduate Fellowship Program, Mrs. Badillo-Urquiola will begin to pursue her Masters of Science in Modeling and Simulation in fall 2014.

**ERIC C. ORTIZ** is a Virtual Learning Environment Developer for the Applied Cognition and Training in Immersive Virtual Environments Lab at the University of Central Florida’s Institute for Simulation and Training. He is currently a student in the Modeling and Simulation Ph.D. program at UCF. Eric is a subject-matter expert in 3D modeling and production pipelines, VE-related software, video production, and interactive technologies.

**IRWIN L. HUDSON** is the Science & Technology Manager responsible for leading the Simulation and Training Technology Center’s Unmanned Ground Systems Research. Mr. Hudson is Contract Officer Representative for several robotic research efforts. He also serves as the Assistant Contract Officer Representative to Dr. Neal Finkelstein for the Research Academic and Operational Support Indefinite Delivery, Indefinite Quantity contract, which supports a large percentage of the research and development budget for STTC’s Blended Simulation Research Branch. He is currently pursuing his Ph.D. in M&S from UCF.